

Data Science Statistics Notes

Median, in statistics, is the middle value of the given list of data when arranged in an order.
Work on ODD Values

Median Find out of Even Values

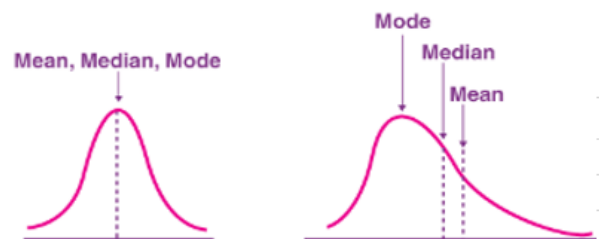
1, 3, 3, **6**, 7, 8, 9

Median = 6

1, 2, 3, **4**, **5**, 6, 8, 9

$$\begin{aligned}\text{Median} &= (4 + 5) \div 2 \\ &= \underline{4.5}\end{aligned}$$

Measures of Central Tendency, Mean, Median & Mode



Students	Heights
Vikas	182
Ashwini	161
Kalpesh	152
Snehal	137
Kumar	121

Students	Heights
Vikas	182
Ashwini	161
Kalpesh	152
Snehal	137

Mean	150.6		Mean	158	AVERAGE(E3:E6)
Median	152		Median	156.5	(161+152)/2
stddev.p	20.77	STDEV.P(B3:B7)			
stddev.s	23.22	STDEV.S(B3:B7)			

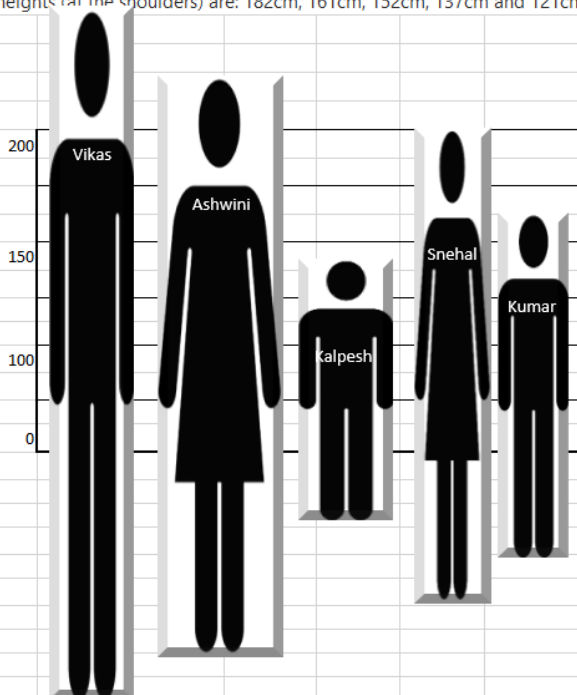
A mode is defined as the value that has a higher frequency in a given set of values. It is the value that appears the most number of times

Students	Heights
Vikas	182
Ashwini	161
Kalpesh	152
Snehal	137
Kumar	182

MODE.SNGL(B16:B20)

Standard Deviation				
	Its symbol is σ (the greek letter sigma)			

The heights (at the shoulders) are: 182cm, 161cm, 152cm, 137cm and 121cm



Students	Heights
Vikas	182
Ashwini	161
Kalpesh	152
Snehal	137
Kumar	121
Average	150.60
STDEV.P	20.771134
STDEV.S	23.222834
STDEVA	23.222834
STDEVPA	20.771134
STDEV	23.222834
STDEVP	20.771134

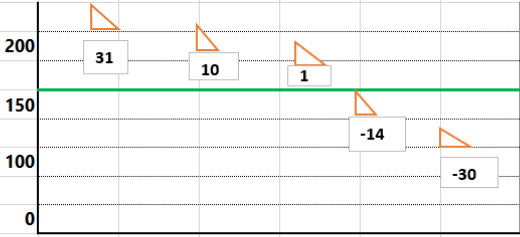
Mean	182	161	152	137	121
	5				
Mean	$\frac{182+161+152+137+121}{5}$				
	150.6				

so the mean (average) height is 151 cm. Let's plot this on the chart:

Now we calculate each student's difference from the Mean

To calculate the Variance, take each difference, square it, and then average the result

Students	Heights	Take each difference
Vikas	182	31
Ashwini	161	10
Kalpesh	152	1
Snehal	137	-14
Kumar	121	-30
Mean	151	



Students	Heights	Take each difference	Square it
Vikas	182	31.40	985.96
Ashwini	161	10	108.16
Kalpesh	152	1	1.96
Snehal	137	-14	184.96
Kumar	121	-30	876.16
Mean	151		2157.2

$$\begin{aligned} &= \frac{(31)^2 + (10)^2 + (1)^2 + (-14)^2 + (-30)^2}{5} \\ &= \frac{985.96 + 108.16 + 1.96 + 184.96 + 876.16}{5} \\ &= \frac{2157.2}{5} = 431.44 \end{aligned}$$

This example is for population

The "Population Standard

But if the data is a sample (a selection taken from a bigger population)
Then the calculation changes

When you have "N" data values that are:

- The population: divide by N when calculating variance (like we did)
- A sample : divide by N-1 when calculating variance

If our 5 student are just a sample of bigger population of student,
We divide by 4 instead of 5 like this

Sample Variance	2157.2 / 4 = 539.3
Sample Standard Deviation	$\sqrt{539.3}$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

Where
 x_i are individual values in the set of data
 \bar{x} is the mean of all x values
 n is the total number of x values in the data set

So the variance is **431.44**
Standard Deviation is just square root of variance, so:

$$\sigma = \sqrt{431.44}$$

STD **20.77**

he good thing about the standard deviation is that it is useful
we can show which heights are within one standard deviation of the mean
7 mm

sing the standard deviation we have a "standard" way of knowing
is normal, and what is extra large or extra small

Population Variance	2157.2/5	431.44
	20.77	

The "Population Standard

The "Sample Standard Deviation"

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

Students	Heights
Vikas	182
Ashwini	161
Kalpesh	152
Snehal	137
Kumar	121

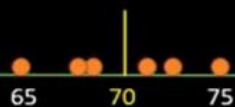
STDEV.P	20.77113
STDEV.S	23.22283
STDEVA	23.22283
STDEVP	20.77113
STDEV	23.22283
STDEVP	20.77113

Looks complicated, but the important change is to
divide by **N-1** (instead of **N**) when calculating a Sample Variance.

History Test

Name	Score
Mohan	75
Andrea	72
Sofia	68
Joe	65
Virat	67
Abdul	73

Average = 70



Math Test

Name	Score
Mohan	93
Andrea	96
Sofia	43
Joe	47
Virat	51
Abdul	90

Average = 70



Average = 70

Name	Score	Abs (Score - Avg)	Abs (Score - Avg)^2
Mohan	75	5	25
Andrea	72	2	4
Sofia	68	2	4
Joe	65	5	25
Virat	67	3	9
Abdul	73	3	9
Avg			12.66
\sqrt{Avg}			3.55

Average = 70

Name	Score	Abs (Score - Avg)	Abs (Score - Avg)^2
Mohan	83	13	169
Andrea	70	0	0
Sofia	70	0	0
Joe	63	7	49
Virat	70	0	0
Abdul	70	0	0
Avg			36.33
\sqrt{Avg}			6.02

Standard Deviation

